Solution to Homework 4

October 23, 2007

1 Problem 1

Given $A \in \mathbb{R}^{m\times n}$ and $B = \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix}$. We assume $\text{rank}(A) = k$. Let $A = U_1 \Sigma V_1^T$ be the thin SVD, where $U_1 \in \mathbb{R}^{m\times k}$, and $V_1 \in \mathbb{R}^{n\times k}$, and $\Sigma \in \mathbb{R}^{k\times k}$.

Let $Bx = \lambda x$ be the eigen-decomposition, where $x \in \mathbb{R}^{(m+n)\times(m+n)}$ is the eigenvector and λ is the corresponding eigenvalue. We further assume $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, where $x_1 \in \mathbb{R}^n$ and $x_2 \in \mathbb{R}^m$. It follows that :

$$Bx = \begin{pmatrix} 0 & V_1 \Sigma U_1^T \\ U_1 \Sigma V_1^T & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} V_1 \Sigma U_1^T x_2 \\ U_1 \Sigma V_1^T x_1 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}.$$

By solving $\lambda x_1 = V_1 \Sigma U_1^T x_2$ and $\lambda x_2 = U_1 \Sigma V_1^T x_1$, we can verify that the eigenvectors and eigenvalues of B can be expressed respectively as

$$x = \begin{pmatrix} \pm V_1 \\ U_1 \end{pmatrix}, \quad \lambda = \pm\ \text{diag}(\Sigma); \quad \text{or} \quad x = \begin{pmatrix} \pm V_2 \\ U_2 \end{pmatrix}, \quad \lambda = 0,$$

where U_2 and V_2 are the null space of U_1 and V_1, respectively.

2 Problem 2

Given $A \in \mathbb{R}^{m\times n}$ and $\text{rank}(A) = r$. Let $A = U\Sigma V^T$ be its SVD, where $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{n\times n}$, and $\Sigma \in \mathbb{R}^{n\times n}$. Note that Σ takes the form of $\Sigma = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r, 0, \cdots, 0)$. It follows that

$$\begin{aligned}
(A^T A + \frac{1}{i} I_n)^{-1} &= \left(V\Sigma U^T U \Sigma V^T + \frac{1}{i} I_n \right)^{-1} \\
&= \left(V \Sigma^2 V^T + \frac{1}{i} I_n \right)^{-1} \\
&= \left(V \tilde{\Sigma}^2 V^T \right)^{-1} \\
&= V \tilde{\Sigma}^{-2} V^T,
\end{aligned}$$

where $\tilde{\Sigma}$ takes the form of $\tilde{\Sigma} = \text{diag}(\sigma_1^2 + 1/i, \cdots, \sigma_r^2 + 1/i, 1/i, \cdots, 1/i) \in \mathbb{R}^{n\times n}$. Moreover, we have

$$\begin{aligned}(A^TA + \frac{1}{i}I_n)^{-1}A^T &= V\tilde{\Sigma}^{-2}V^TV\Sigma U^T \\ &= V\tilde{\Sigma}^{-2}\Sigma U^T \\ &= V\hat{\Sigma}U^T,\end{aligned}$$

where $\hat{\Sigma}$ takes the form of $\hat{\Sigma} = \text{diag}(\frac{\sigma_1}{\sigma_1^2+1/i}, \cdots, \frac{\sigma_r}{\sigma_r^2+1/i}, 0, \cdots, 0) \in \mathbb{R}^{n\times n}$.

Let $A = U_1\text{diag}(\sigma_1, \cdots, \sigma_r)V_1^T$ be the thin SVD, where $U_1 \in \mathbb{R}^{m\times r}$ and $V_1 \in \mathbb{R}^{r\times r}$. Thus we have $A^+ = V_1\text{diag}(1/\sigma_1, \cdots, 1/\sigma_r)U_1^T$. It follows that

$$\begin{aligned}B_i - A^+ &= V\hat{\Sigma}U^T - V_1\text{diag}(1/\sigma_1, \cdots, 1/\sigma_r)U_1^T \\ &= V_1\left(\text{diag}(\frac{\sigma_1}{\sigma_1^2+1/i}, \cdots, \frac{\sigma_r}{\sigma_r^2+1/i}) - \text{diag}(1/\sigma_1, \cdots, 1/\sigma_r)\right)U_1^T \\ &= (-V_1)\ \text{diag}\left(\frac{1}{\sigma_1(i\sigma_1^2+1)}, \cdots, \frac{1}{\sigma_r(i\sigma_r^2+1)}\right)\ U_1^T.\end{aligned}$$

It can then be verified $\max_j\left(\frac{1}{\sigma_j(i\sigma_j^2+1)}\right) = \frac{1}{\sigma_r(i\sigma_r^2+1)}$, which implies that

$$\|B_i - A^+\|_2 = \frac{1}{\sigma_r(A)(i\sigma_r(A)^2+1)}.$$

Recall that B_i takes the form :

$$(A^TA + \frac{1}{i}I_n)^{-1}A^T = V\hat{\Sigma}U^T,$$

where $\hat{\Sigma}$ takes the form of $\hat{\Sigma} = \text{diag}(\frac{\sigma_1}{\sigma_1^2+1/i}, \cdots, \frac{\sigma_r}{\sigma_r^2+1/i}, 0, \cdots, 0)$. An obvious result is : as $i \to \infty$, we have $\hat{\Sigma} \to \text{diag}(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_r}, 0, \cdots, 0)$, and $B_i = V\hat{\Sigma}U^T \to A^+ = V_1\Sigma^{-1}U_1^T$. This completes the proof.

3 Problem 3

Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, let $A = U\Sigma U^T$ be its SVD, where $U \in \mathbb{R}^{n\times n}$ is orthogonal and $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_n)$. It follows that

$$\begin{aligned}\text{trace}(G^TAG) &= \text{trace}\left(G^TU\Sigma U^TG\right) \\ &= \text{trace}\left((U^TG)^T\Sigma(U^TG)\right) \\ &= \text{trace}\left(M^T\Sigma M\right),\end{aligned}$$

where $M = U^TG \in \mathbb{R}^{n\times m}$. Note that M is column-orthogonal. Let $N \in \mathbb{R}^{n\times(n-m)}$ be the null space of M. We have

$$\begin{aligned}\text{trace}([M,N]^T[M,N]) &= \text{trace}(I_n) \\ &= \text{trace}([M,N][M,N]^T) \\ &= \text{trace}(MM^T) + \text{trace}(NN^T).\end{aligned}$$

Denote the diagonal entries of MM^T and NN^T as M_{ii} and N_{ii}, where $MM^T, NN^T \in \mathbb{R}^{n\times n}$. It can be easily verified that $M_{ii} + N_{ii} = 1$, and $M_{ii}, N_{ii} >= 0$. Thus we have $0 \leq M_{ii} \leq 1$ for $i = 1, \cdots, n$. Furthermore, using the result $\text{trace}(MM^T) = \text{trace}(M^T M) = m$, we have $\sum_{i=1}^n M_{ii} = m$.

Recall that

$$\begin{aligned}
\text{trace}(G^T AG) &= \text{trace}\left(M^T \Sigma M\right) \\
&= \text{trace}\left(\Sigma M M^T\right) \\
&= \sum_{i=1}^{n} (\sigma_i M_{ii}) \\
&\leq \sum_{i=1}^{m} \sigma_i.
\end{aligned}$$

It can be easily verified that when $G = [u_1, u_2, \cdots, u_m]$, we have $\text{trace}(G^T AG) = \text{trace}(U_1^T U \Sigma U^T U_1) = \sum_{i=1}^m \sigma_i$. This completes the proof.